Nonparametric estimation of regression level sets
Thomas Laloë, Rémi Servien

To cite this version:
Thomas Laloë, Rémi Servien. Nonparametric estimation of regression level sets. 2011. hal-00674197v2

HAL Id: hal-00674197
https://hal.archives-ouvertes.fr/hal-00674197v2
Submitted on 15 Mar 2012 (v2), last revised 8 Oct 2012 (v3)
Nonparametric estimation of regression level sets

T. Laloë\textsuperscript{a,}\textsuperscript{*}, R. Servien\textsuperscript{b}

\textsuperscript{a}Université de Nice Sophia Antipolis, Laboratoire J.-A. Dieudonné, Parc Valrose, 06108 Nice Cedex 2, France, Tel.: +33492076020.
\textsuperscript{b}Montpellier SupAgro-INRA, UMR MISTEA 729, 2 place Pierre Viala, 34060 Montpellier Cedex 2, France.

Abstract

Let \((X, Y)\) be a random pair taking values in \(\mathbb{R}^d \times J\), where \(J \subset \mathbb{R}\) is supposed to be bounded. We propose a plug-in estimator of the level sets of a regression function \(r\) of \(Y\) on \(X\), using a kernel estimator of \(r\). We consider an error criterion defined by the volume of the symmetrical difference between the real and estimated level sets. We state the consistency of our estimator, and we get a rate of convergence equivalent to the one obtained by Cadre (2006) for the density function level sets.

Keywords: Regression function, Level set, Plug-in estimator, Kernel estimator.

1. Introduction

In this paper, we consider the problem of estimating the level sets of a regression function. More precisely, consider a random pair \((X, Y)\) taking values in \(\mathbb{R}^d \times J\), where \(J \subset \mathbb{R}\) is supposed to be bounded. The goal of this paper is then to build an estimator of the level sets of the regression function \(r\) of \(Y\) on \(X\), defined for all \(x \in \mathbb{R}^d\) by

\[
r(x) = \mathbb{E}[Y|X = x].
\]

For \(t > 0\), a level set for \(r\) is defined by

\[
\mathcal{L}(t) = \{x \in \mathbb{R}^d : r(x) > t\}.
\]

Assume that we have an independent and identically distributed sample (i.i.d.) \(\left((X_1, Y_1), \ldots, (X_n, Y_n)\right)\) with the same distribution than \((X, Y)\). We then consider a plug-in estimator of \(\mathcal{L}(t)\). More precisely, we use a consistent estimator \(\hat{r}_n\) of \(r\), in order to estimate \(\mathcal{L}(t)\) by

\[
\mathcal{L}_n(t) = \{x \in \mathbb{R}^d : \hat{r}_n(x) > t\}.
\]
Most of the research works on the estimation of level sets concern the density function. One can cite the works of Cadre [1], Cuevas and Fraiman [2], Hartigan [3], Polonik [4], Tsybakov [5], Walther [6]. This large number of works on this subject is motivated by the high number of possible applications. Estimating these level sets can be useful in mode estimation (Müller and Stawitzki [7], Polonik [4]), or in clustering (Biau, Cadre and Pelletier [8], Cuevas, Febrero and Fraiman [9, 10]). In particular, Biau, Cadre and Pelletier [8] use an estimator of the level sets of the density function to determine the number of clusters.

The same applications are possible with the regression function. Moreover, it is for instance possible to use an estimator of the level sets of the regression function to determine the path of water flow from a digital representation of an area. In the same vein, in medical imaging, people want to estimate the areas where some function of the image exceeds a fixed threshold. It may be useful, for instance in order to automatically determine the location or the nature of a tumor. Note that, in these two examples, the use of a compact set $J$ is fully justified. This is generally the case in most practical situations, particularly in image analysis.

Despite the many potential applications, the estimation of the level sets of the regression function has not been widely studied. Müller [11] mentioned it briefly in his survey. One can also cite the recent work of Cavalier [12], Polonik and Wang [13], Scott and Davenport [14], and Willett and Nowak [15]. The estimator proposed by Cavalier is based on the maximization of the excess mass. It is an adaptation of the estimator proposed by Tsybakov [5] for the density function. Scott and Davenport use a cost sensitive approach. The main advantage of our estimator is the simplicity of his calculation. Moreover, unlike the estimator proposed by Cavalier, our estimator does not require that the level sets are star shaped.

All our consistency results are in the sense of the symmetrical difference (Figure 1), defined by
\[ \mathcal{L}_n \Delta \mathcal{L} = (\mathcal{L}_n \cap \mathcal{L}^C) \cup (\mathcal{L}_n^C \cap \mathcal{L}), \]
where $\mathcal{L}_n = \mathcal{L}_n(t)$ and $\mathcal{L} = \mathcal{L}(t)$.

Our goal is to establish some consistency results under reasonable assumptions on $r$ and $\hat{r}_n$. We first use the results by Cuevas, González-Manteiga and Rodríguez-Casal [16] in order to state a consistency result for any consistent estimator $\hat{r}_n$ of $r$. Then we particularize our approach by considering a kernel estimator of the regression function. For this estimator, we get a rate of convergence equivalent to the one obtained by Cadre [1] for the density function.

This paper is organized as follows. We give the main results in Section 2, and proofs are collected in Section 3.
2. Main results

2.1. Consistency

From now on, $\|\cdot\|$ stands for the Euclidean norm on a finite dimensional space. Besides, for all integrable function $g : \mathbb{R}^d \to \mathbb{R}$, we denote by $\|g\|_p$ the norm defined by

$$\|g\|_p = \left( \int_{\mathbb{R}^d} |g(x)|^p dx \right)^{1/p}.$$ 

2.1.1. General estimator $\hat{r}_n$ of $r$

In this paragraph, we assume that we know a consistent estimator $\hat{r}_n$ of $r$ (in a sense defined hereafter). Let us introduce the following assumption on the regression function $r$:

**A0** There exists $t^- < t$ such that $\mathcal{L}(t^-)$ is compact. Besides, $\lambda(\{r = t\}) = 0$ (where $\lambda$ stands for the Lebesgue measure).

Roughly speaking, the last part of this assumption means that the Lebesgue measure does not charge the set $\{x \in \mathbb{R}^d : r(x) = t\}$.

Theorem 3 by Cuevas, González-Manteiga and Rodríguez-Casal [16] states the almost sure consistency of $\lambda(\mathcal{L}_n \Delta \mathcal{L})$ for any consistent estimator $\hat{r}_n$ of $r$. Theorem 2.1 below deals with the consistency of $\mathbb{E}\lambda(\mathcal{L}_n \Delta \mathcal{L})$. 

Figure 1: Symmetrical difference (in black) between two sets $A$ (in red) and $B$ (in blue).
Theorem 2.1. Under Assumption A0, if
\[ \sup_{\mathbb{R}^d} |\hat{r}_n - r| \to 0 \quad \text{a.s.}, \]
then
\[ E \lambda \left( \mathcal{L}_n(t) \Delta \mathcal{L}(t) \right) \to 0. \]

Then, provided that \( \hat{r}_n \) is consistent with respect to the \( L_p \) or supremum norm, the plug-in estimator of the level sets of the regression function is consistent with respect to the symmetrical difference. For example, the BSE estimator (Bosq and Lecoutre [17], Chapter 7), the \( k \)-nearest neighbor estimator ([17] Chapter 8), as well as the regressogram ([17], Chapter 6) satisfy this property. In the next paragraph, we focus on the case of the kernel estimator.

2.1.2. Kernel estimator \( r_n \) of \( r \)
Let us consider now the case of the kernel estimator of the regression function. Assume that we can write
\[ r(x) = \frac{\varphi(x)}{f(x)}, \]
where \( f \) is the density function of \( X \), and \( \varphi \) is defined by \( \varphi(x) = r(x)f(x) \).

Let \( K \) be a kernel on \( \mathbb{R}^d \), that is a probability density on \( \mathbb{R}^d \). We denote \( K_h(x) = K(x/h) \). From an i.i.d. sample \( \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \), we define, for all \( x \in \mathbb{R}^d \),
\[ \varphi_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} Y_i K_h(x - X_i) \quad \text{and} \quad f_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K_h(x - X_i). \]

For all \( x \in \mathbb{R}^d \), the kernel estimator of \( r \) is then defined by
\[ r_n(x) = \begin{cases} \frac{\varphi_n(x)}{f_n(x)} & \text{if } f_n(x) \neq 0 \\ 0 & \text{otherwise}. \end{cases} \]

For a complete study of the kernel estimator of the regression function and its properties, we refer to the book of Prakasa Rao [18].

We derive the following result from Theorem 2.1 using the consistency properties of the kernel estimator of the regression function (Bosq and Lecoutre [17]).

Corollary 2.1. Under Assumption A0, if \( K \) is bounded, integrable, with compact support and Lipschitz, and if \( h \to 0 \) and \( nh^d / \log n \to \infty \), then
\[ E \lambda \left( \mathcal{L}_n(t) \Delta \mathcal{L}(t) \right) \to 0. \]

This result gives concrete conditions to obtain the consistency of \( \mathcal{L}_n(t) \Delta \mathcal{L}(t) \) for the kernel estimator. It then particularizes the previous result from Cuevas, González-Manteiga and Rodríguez-Casal [16]. In the next paragraph, we establish a rate of convergence for this estimator.
2.2. Rate of convergence

Remember that \( r_n \) is the kernel estimator of the regression function. From now on, \( \Theta \subset (0, \sup_{\mathbb{R}^d} r) \) is an open interval. Let us introduce the following assumptions:

**A1** The functions \( r \) and \( f \) are twice continuously differentiable, and, \( \forall t \in \Theta, \exists 0 < t^- < t : \inf_{\mathcal{L}(t^-)} f > 0; \)

**A2** For all \( t \in \Theta, \)

\[
\inf_{r^{-1}(\{t\})} \|\nabla r\| > 0,
\]

where, \( \nabla \psi(x) \) stands for the gradient at \( x \in \mathbb{R}^d \) of the differentiable function \( \psi : \mathbb{R}^d \to \mathbb{R}. \)

Let us mention that under Assumptions A1 and A2, we have (Proposition A.2 in [1])

\[
\forall t \in \Theta : \lambda(r^{-1}[t - \varepsilon, t + \varepsilon]) \to 0 \quad \text{as} \quad \varepsilon \to 0.
\]

This property, first used by Polonik [4], is almost identical to the last part of Assumption A0 in Paragraph 2.1.1. Let us now introduce the assumptions on the kernel \( K. \)

**A3** \( K \) is a continuously differentiable with a compact support. Moreover, there exists a decreasing function \( \mu : \mathbb{R}^+ \to \mathbb{R} \) such that \( K(x) = \mu(\|x\|) \) for all \( x \in \mathbb{R}^d. \)

From now on, we denote by \( \partial A \) the boundary of any subset \( A \subset \mathbb{R}^d. \) Besides, we introduce \( \mathcal{H} \) the \((d - 1)\)-dimensional Hausdorff measure (Evans and Gariepy [19]). Recall that \( \mathcal{H} \) agrees with ordinary “(k - 1)-dimensional surface area” on nice sets (Proposition A.1 in [1]). Finally, we set \( \tilde{K} = \int K^2 d\lambda. \)

We are now in a position to establish a rate of convergence for \( \mathbb{E} \lambda(\mathcal{L}_n(t)\Delta \mathcal{L}(t)). \)

**Theorem 2.2.** Under Assumptions A0 – A3, if \( nh^d/(\log n)^7 \to \infty \) and \( nh^{d+4}\log n \to 0, \) then

\[
\mathbb{E} \lambda(\mathcal{L}_n(t)\Delta \mathcal{L}(t)) = O(1/\sqrt{nh^d}).
\]

Note that our estimator is easy to calculate and requires reasonable assumptions. Cavalier [12] gets a better rate, but with an estimator more difficult to calculate and with more restrictive assumptions. For example, we do not need for the level sets to be star-shaped. However, an interesting problem left for future discussion is the choice of the bandwidth \( h \) in our estimator. Indeed, an optimal bandwidth for estimating \( r \) is not necessarily optimal for estimating \( \mathcal{L}. \)
3. Proofs

This section is dedicated to the proofs of Theorem 2.1 and Theorem 2.2. From now on, \( c, C_1, \) and \( C_2 \) are non-negative constants, whose values may change from line to line.

3.1. Proof of Theorem 2.1

Let \( \varepsilon > 0 \), for all \( n \in \mathbb{N} \), we define \( E_{n,\varepsilon} = \{ x \in \mathcal{L}_n(t) \cup \mathcal{L}(t) : |\hat{r}_n(x) - r(x)| \leq \varepsilon \} \). Since \( \sup_{d} |\hat{r}_n - r| \to 0 \) a.s., we have that \( \mathcal{L}_n(t) \subset \mathcal{L}(t^-) \) for \( n \) large enough which gives us \( E_{n,\varepsilon} = \{ x \in \mathcal{L}(t^-) : |\hat{r}_n(x) - r(x)| \leq \varepsilon \} \).

Moreover we have

\[
\lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t)) = \lambda((\mathcal{L}_n(t) \Delta \mathcal{L}(t)) \cap E_{n,\varepsilon}) + \lambda((\mathcal{L}_n(t) \Delta \mathcal{L}(t)) \cap E_{n,\varepsilon}^c).
\]

Since \( (\mathcal{L}_n(t) \Delta \mathcal{L}(t)) \cap E_{n,\varepsilon} \subset \{ t - \varepsilon \leq r \leq t + \varepsilon \} \), we obtain

\[
\lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t)) \leq \lambda(\{ t - \varepsilon \leq r \leq t + \varepsilon \}) + \lambda(E_{n,\varepsilon}^c),
\]

and

\[
\mathbb{E}\lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t)) \leq \lambda(\{ t - \varepsilon \leq r \leq t + \varepsilon \}) + \mathbb{E}\lambda(E_{n,\varepsilon}^c).
\]

Finally, as \( \sup_{d} |\hat{r}_n - r| \to 0 \) a.s., then \( \lambda(E_{n,\varepsilon}^c) \to 0 \) and, according to assumption \( A0 \), \( \mathbb{E}\lambda(E_{n,\varepsilon}^c) \to 0 \) by the Lebesgue’s dominated convergence theorem.

3.2. Proof of Theorem 2.2

In this proof, some arguments are classical result from the kernel density (or regression) estimation theory. For more details, we refer the reader to the book by Bosq and Lecoutre ([17]), chapter 4 and 5.

3.2.1. Preliminary results

All the results in this sections are stated under Assumptions \( A0 - A3 \). The proof of the theorem relies on the four following lemmas.

Let us define

\[
\Omega_{n,c} = \left\{ \sqrt{nh^d} \sup_{\mathcal{L}_n(t) \cup \mathcal{L}(t)} |r_n - r| \geq c\sqrt{\log n} \right\}.
\]

Lemma 3.1. If \( nh^{d+4}/\log n \to 0 \), then there exists \( \Gamma > 0 \) such that

\[
\sqrt{nh^d}\mathbb{P}(\Omega_{n,\Gamma}) \to 0.
\]
Note that the condition $nh^{d+4}/\log n \to 0$ is satisfied under the assumptions of Theorem 2.2.

**Proof of Lemma 3.1**

As $r$ is continuous, we have $\sup_{\mathcal{L}(t^-)} |r| < c$. Assuming that $\inf_{\mathcal{L}(t^-)} f > 0$, then, since $\sup_{\mathcal{L}(t^-)} |f_n - f| \to 0$ a.s. under the assumptions of Lemma 3.1 (Bosq and Lecoutre [17]), there exists $\theta > 0$ such that $\inf_{\mathcal{L}(t^-)} f_n > \theta$ a.s. for $n$ large enough. So we can write

$$\sup_{\mathcal{L}(t^-)} |r_n - r| = \sup_{\mathcal{L}(t^-)} \left| \frac{\varphi_n - \varphi}{f_n} + r \frac{f_n - f}{f_n} \right| \leq c \left( \sup_{\mathcal{L}(t^-)} |\varphi_n - \varphi| + \sup_{\mathcal{L}(t^-)} |f_n - f| \right).$$

(2)

We have

$$\sup_{\mathcal{L}(t^-)} |\varphi_n - \varphi| \leq \sup_{\mathcal{L}(t^-)} |\varphi_n - \mathbb{E}\varphi_n| + \sup_{\mathcal{L}(t^-)} |\mathbb{E}\varphi_n - \varphi|.$$

We cover $\mathcal{L}(t^-)$ with $\ell_n$ balls $B_k = B(x_k, \rho_n)$ $(k = 1, \ldots, \ell_n)$ of radius $\rho_n$.

Consider $x \in \mathcal{L}(t^-)$, we denote by $B_k$ the ball containing $x$. Then we set, for $x, x' \in \mathcal{L}(t^-)$,

$$A_n(x, x') = \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ K_h(x - X_i) - K_h(x' - X_i) \right]$$

$$- \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ K_h(x - X_i) - K_h(x' - X_i) \right],$$

which leads us to

$$\sup_{\mathcal{L}(t^-)} |\varphi_n - \varphi| \leq \sup_{1 \leq k \leq \ell_n} |\varphi_n(x_k) - \mathbb{E}\varphi_n(x_k)| + \sup_{x \in \mathcal{L}(t^-)} |A_n(x, x_k)| + \sup_{\mathcal{L}(t^-)} |\mathbb{E}\varphi_n - \varphi|. $$

(3)

Then, since $K$ is Lipschitz, there exists $\gamma > 0$ such that

$$|A_n(x, x_k)| \leq c h^{-d-\gamma} \rho_n^\gamma \left( \frac{1}{n} \sum_{i=1}^{n} |Y_i| + \mathbb{E}|Y| \right)$$

$$\leq c h^{-d-\gamma} \rho_n^\gamma$$

since $Y$ is bounded.

As a consequence, we have

$$\mathbb{P} \left( \sup_{x \in \mathcal{L}(t^-)} |A_n(x, x_k)| > \frac{c}{4} \sqrt{\frac{\log n}{nh^d}} \right) \leq \mathbb{P} \left( ch^{-d-\gamma} \rho_n^\gamma > \frac{c}{4} \sqrt{\frac{\log n}{nh^d}} \right).$$

One can choose

$$\rho_n = n^{-a}, a > 0 \quad \text{and} \quad \rho_n^\gamma = o \left( h^{d+\gamma} \sqrt{\frac{\log n}{nh^d}} \right),$$
such that
\[
P \left( \sup_{x \in \mathcal{L}(t^-)} |A_n(x, x_k)| > \log n / \sqrt{nh^d} \right) = 0. \tag{4}
\]

Then, using the arguments of the proof of Theorem 5.II.3 in [17], we obtain
\[
\forall \varepsilon > 0, \ P \left( \sup_{1 \leq k \leq l_n} |\varphi_n(x_k) - E\varphi_n(x_k)| > \varepsilon \right) < 2\ell_n e^{-\frac{nh^d}{c}}.
\]

If we set \( \varepsilon = \varepsilon_0 \sqrt{\log n / nh^d} \), we have
\[
P \left( \sup_{1 \leq k \leq l_n} |\varphi_n(x_k) - E\varphi_n(x_k)| > \varepsilon_0 \sqrt{\log n / nh^d} \right) \leq c\ell_n n^{-2\varepsilon_0/c} \leq cn^{-2\varepsilon_0/c} \rho_n^{-d}. \tag{5}
\]

Remember that \( \rho_n = n^{-a} \), with \( a > 0 \), one gets
\[
\sqrt{nh^d} P \left( \sup_{1 \leq k \leq l_n} |\varphi_n(x_k) - E\varphi_n(x_k)| \geq c \sqrt{\log n / nh^d} \right) \rightarrow 0.
\]

which tends to 0 choosing \( \varepsilon_0 > (1/2 + ad)c \).

Moreover, under \( \mathbf{A3} \), \( K \) is even which gives us
\[
\sup_{\mathcal{L}(t^-)} |E \varphi_n - \varphi| = O \left( \sqrt{\log n / nh^d} \right),
\]

and, using that \( nh^{d+4} / \log n \to 0 \) we obtain
\[
\sqrt{nh^d} P \left( \sup_{\mathcal{L}(t^-)} |E \varphi_n - \varphi| \geq c \sqrt{\log n / nh^d} \right) \rightarrow 0. \tag{6}
\]

From (3) and using (4), (5) and (6) we obtain
\[
\sqrt{nh^d} P \left( \sup_{\mathcal{L}(t^-)} |\varphi_n - \varphi| \geq c \sqrt{\log n / nh^d} \right) \rightarrow 0.
\]

From (2) and such as \( \sup_{\mathcal{L}(t^-)} |f_n - f| \to 0 \) a.s., we conclude the proof. \( \square \)

Consider \( t \in \Theta \). For all \( x \in \mathcal{L}(t^-) \), we define
\[
V_n(x, t) = \text{Var}(Y - t)K_h(x - X) \quad \text{and} \quad \tilde{E}r_n(x) = E\varphi_n(x) / E f_n(x).
\]
For all $x \in \mathcal{L}(t^-)$ such that $V_n(x,t) \neq 0$, we set

$$t_n(x) = \mathbb{E} f_n(x) \sqrt{\frac{n h^d}{V_n(x,t)}} (t - \mathbb{E} r_n(x)).$$

Besides, we consider the sets

$$\mathcal{V}_n^t = r^{-1}[t, t + \Gamma \sqrt{\log n / nh^d}] \cap \mathcal{L}(t^-)$$

and

$$\widetilde{\mathcal{V}}_n^t = r^{-1}[t - \Gamma \sqrt{\log n / nh^d}, t] \cap \mathcal{L}(t^-).$$

Finally, we denote by $\Phi$ the distribution function of the standard normal $\mathcal{N}(0,1)$, and we define $\overline{\Phi}(x) = 1 - \Phi(x)$.

**Lemma 3.2.** There exists $c > 0$ such that for all $n \geq 1, t \in \mathbb{R}$ and $x \in \mathcal{L}(t^-)$:

$$|\mathbb{P}(r_n(x) \leq t) - \Phi(t_n(x))| \leq \frac{c}{\sqrt{nh^d}}.$$

**Proof of Lemma 3.2**

Set, for $i = 1, \ldots, n$,

$$Z_i(x,t) = (Y_i - t) K_h(x - X_i), \quad Z(x,t) = (Y - t) K_h(x - X).$$

By definition, we have $V_n(x,t) = \text{Var}(Z(x,t))$, and

$$\begin{align*}
\mathbb{P}(r_n(x) < t) &= \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} Z_i(x,t) < 0\right) \\
&= \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} \left(Z_i(x,t) - \mathbb{E} Z(x,t)\right) < -\mathbb{E} Z(x,t)\right) \\
&= \mathbb{P}\left(\frac{1}{\sqrt{V_n(x,t)}} \frac{1}{n} \sum_{i=1}^{n} \left(Z_i(x,t) - \mathbb{E} Z(x,t)\right) < t_n(x)\right).
\end{align*}$$

Then, the Berry-Essen inequality gives us

$$|\mathbb{P}(r_n(x) < t) - \Phi(t_n(x))| \leq \frac{c}{\sqrt{V_n(x,t)}} \mathbb{E} |(Y - t) K_h(x - X) - \mathbb{E} (Y - t) K_h(x - X)|^3.$$

(7)

Finally, under Assumptions A1 and A3, we have (see for example Bosq and Lecoutre [17])

$$\sup_{x \in \mathcal{L}(t^-)} |(Y - t) K_h(x - X) - \mathbb{E} (Y - t) K_h(x - X)|^3 \leq c h^d$$

and

$$\inf_{x \in \mathcal{L}(t^-)} V_n(x,t) \geq c h^d.$$
The lemma can then be deduced from (7). □

Define now \( \Theta_0 \) the set of all \( t \) in \( \Theta \) such that

\[
\lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon} \lambda\left(r^{-1}[t - \varepsilon, t]\right) = \lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon} \lambda\left(r^{-1}[t, t + \varepsilon]\right) = \int_{\partial \mathcal{L}(t)} \|\nabla r\|^{-1} d\mathcal{H}.
\]

The following result is proven in Cadre [1] (Lemma 3.2).

**Lemma 3.3.** \( \Theta_0 = \Theta \) almost everywhere.

Note that under Assumptions \( A1 \) and \( A2 \), we obtain, thanks to Proposition A.2 in [1],

\[
\lambda\left(r^{-1}[t - \varepsilon, t + \varepsilon]\right) = \lambda\left(r^{-1}(t - \varepsilon, t + \varepsilon)\right),
\]

for all \( t \in \Theta \) and \( \varepsilon > 0 \) small enough.

Finally, we set

\[
v(x) = \text{Var}(Y|X = x) + r^2(x),
\]

and, for \( t \in \Theta \) and \( x \in \mathcal{L}(t^-)\),

\[
\tilde{t}_n(x) = f(x) \sqrt{\frac{nh^d}{K f(x)(v(x) + t^2)}} (t - r(x)).
\]

We are now in a position to prove Lemma 3.4 below.

**Lemma 3.4.** If \( nh^d/\log n \to \infty \) and \( nh^{d+4} \log n \to 0 \), then for all \( t \in \Theta_0 \),

\[
\lim_{n \to \infty} \sqrt{nh^d} \left[ \int_{\mathcal{V}_n^t} \mathbb{P}(r_n(x) < t) dx - \int_{\mathcal{V}_n^t} \Phi(\tilde{t}_n(x)) dx \right] = 0,
\]

and

\[
\lim_{n \to \infty} \sqrt{nh^d} \left[ \int_{\mathcal{V}_n^t} \mathbb{P}(r_n(x) > t) dx - \int_{\mathcal{V}_n^t} \Phi(\tilde{t}_n(x)) dx \right] = 0.
\]

**Proof of Lemma 3.4** We only prove the first equation, the second one can be obtained with similar arguments.

Define \( E_n \) by

\[
E_n = \sqrt{nh^d} \int_{\mathcal{V}_n^t} |\Phi(t_n(x)) dx - \Phi(\tilde{t}_n(x)) dx|.
\]

As \( \Phi \) is Lipschitz we have

\[
E_n \leq c\sqrt{nh^d} \lambda(V_n^t) \sup_{V_n^t} |t_n - \tilde{t}_n|.
\]

By definition of \( t_n(x) \) and \( \tilde{t}_n(x) \), we have, for all \( x \in V_n^t \),
\[
\frac{1}{\sqrt{n}h^d}|t_n(x) - \tau_n(x)|
\leq |t - r(x)| \left| \frac{f(x)}{\sqrt{Kf(x)(v(x) + t^2)}} - \frac{\mathbb{E} f_n(x)}{\sqrt{V_n(x, t)h^d}} \right|
\]
\[
+ \sqrt{\frac{h^d}{V_n(x, t)}} \mathbb{E} f_n(x) |r(x) - \mathbb{E} r_n(x)|
\leq \sqrt{\log n} \frac{n}{nh^d} \left| \frac{f(x)V_n(x, t)h^{-d} - (\mathbb{E} f_n(x))^2 \tilde{K}(v(x) + t^2)}{\tilde{K}(v(x) + t^2)V_n(x, t)h^{-d}} \right|
\]
\[
+ \sqrt{\frac{h^d}{V_n(x, t)}} \mathbb{E} f_n(x) |r(x) - \mathbb{E} r_n(x)| \tag{9}
\]
Remember that
\[
|\mathbb{E} r_n - r| \leq \frac{1}{f_n} |\mathbb{E}\phi_n - \phi| + |r| |\mathbb{E}|f_n - f| \tag{10}
\]
Since \(V_n\) is included in \(\mathcal{L}(t^-)\), we can deduce (Bosq and Lecoutre [17]) from A1, A3 and (10) that
\[
\sup_{V_n} |\mathbb{E} r_n - r| \leq c h^2. \tag{11}
\]
Moreover, if we set
\[
V_n^1(x) = \text{Var } K_h(x - X), \quad V_n^2 = \text{Var } Y K_h(x - X),
\]
we can write
\[
|f(x)V_n(x, t)h^{-d} - (\mathbb{E} f_n(x))^2 \tilde{K}(v(x) + t^2)|
\leq |f(x)| \left| V_n(x, t)h^{-d} - \tilde{K}\mathbb{E} f_n(x)(v(x) + t^2) \right| + c|f(x) - \mathbb{E} f_n(x)|
\leq |f(x)| \left| V_n(x, t)h^{-d} - \tilde{K}f(x)(v(x) + t^2) \right| + c|f(x) - \mathbb{E} f_n(x)|
\leq |f(x)| \left( |V_n^1(x)h^{-d} - \tilde{K}f(x)| + |V_n^2(x)h^{-d} - \tilde{K}f(x)v(x)| \right.
\quad + 2t |\text{Cov}(Y K_h(x - X), K_h(x - X))| \bigg) + c|f(x) - \mathbb{E} f_n(x)|
\leq c \left( |V_n^1(x)h^{-d} - \tilde{K}f(x)| + |V_n^2(x)h^{-d} - \tilde{K}f(x)v(x)| \right.
\quad + |\text{Cov}(Y K_h(x - X), K_h(x - X))| + |f(x) - \mathbb{E} f_n(x)| \bigg).
\]
Again, since \(V_n \subset \mathcal{L}(t^-)\), we can deduce (Bosq and Lecoutre [17]) from A1 and A3 that
\[
\sup_{x \in V_n} |f(x)V_n(x, t)h^{-d} - (\mathbb{E} f_n(x))^2 \tilde{K}(v(x) + t^2)| \leq c h. \tag{12}
\]
We deduce from (9), (11) and (12) that
\[ \sup_{x \in V_n} |t_n(x) - \tilde{t}_n(x)| \leq c \left( \sqrt{h \log n} + \sqrt{n h^{k+4}} \right). \]

Then, thanks to (8) and since \( t \in \Theta_0 \), we have for \( n \) large enough
\[ E_n \leq c \sqrt{\log n} \left( \sqrt{h \log n} + \sqrt{n h^{k+4}} \right). \]

Finally, Lemma 3.2 leads us to
\[ \sqrt{n h^d} \left[ \int_{V_n^t} \mathbb{P}(r_n(x) < t) dx - \int_{V_n^t} \Phi(t_n(x)) dx \right] \leq c \lambda(V_n^t) \]
which tends to 0 since \( \lambda(r^{-1}[t - \varepsilon, t + \varepsilon]) \to 0 \). This and (13) ends the proof. □

3.2.2. Proof of Theorem 2.2

We first note that
\[ \mathbb{E} \left( \mathcal{L}(t) \Delta \mathcal{L}(t) \right) = \int_{\mathcal{L}(t)^c \cap \{r \geq t\}} \mathbb{P}(r_n(x) < t) dx + \int_{\mathcal{L}(t)^c \cap \{r < t\}} \mathbb{P}(r_n(x) \geq t) dx. \]

Set
\[ \mathbb{P}_{n,t}(x) = \mathbb{P}(r_n(x) < t), \quad \overline{\mathbb{P}}_{n,t}(x) = \mathbb{P}(r_n(x) \geq t) \]
and remember that
\[ V_n^t = r^{-1}[t, t + \Gamma \sqrt{\log n/n h^d}] \cap \mathcal{L}(t^-) \quad \text{and} \quad \overline{V}_n^t = r^{-1}[t - \Gamma \sqrt{\log n/n h^d}, t] \cap \mathcal{L}(t^-). \]

Consider \( t \in \Theta_0 \) and define
\[ I_n = \int_{V_n^t} \Phi(\tilde{t}_n(x)) dx, \quad \overline{I}_n = \int_{\overline{V}_n^t} \overline{\Phi}(\tilde{t}_n(x)) dx. \]

We have
\[ I_n = \frac{1}{\sqrt{2 \pi K}} \int_{V_n^t} \int_{-\infty}^{b_n(x)} \exp \left( -\frac{u^2}{2K} \right) du dx \]
where \( b_n(x) = \sqrt{f(x) n h^d (t - r(x)) / \sqrt{v(x) + t^2}} \).

Besides, \( b_n(x) = \sqrt{|\varphi(x)| / v(x) + t^2} b_n'(x) \),
with \( b_n'(x) = \sqrt{n h^d (t - r(x)) / \sqrt{|r(x)|}} \). Then we can find two positive constants \( C_1 \) and \( C_2 \) such that
\[ C_1 b_n'(x) \leq b_n(x) \leq C_2 b_n'(x), \]
which leads us to
\[ I_n \geq \frac{C_1}{\sqrt{2\pi K}} \int_{V_n} \int_{-\infty}^{b_n(x)} \exp \left( -\frac{C_1^2 u^2}{2K} \right) \, du \, dx, \]
and
\[ I_n \leq \frac{C_2}{\sqrt{2\pi K}} \int_{V_n} \int_{-\infty}^{b_n(x)} \exp \left( -\frac{C_2^2 u^2}{2K} \right) \, du \, dx. \]

Using the arguments of the proof of Proposition 3.1 in [1], we obtain
\[ C_1 \frac{\sqrt{tK}}{\sqrt{2\pi}} \int_{\partial L(t)} \frac{1}{\| \nabla r \|} \, d\mathcal{H} \leq \lim_{n \to \infty} \sqrt{nh^d} I_n \leq \lim_{n \to \infty} \sqrt{nh^d} I_n \leq C_2 \frac{\sqrt{tK}}{\sqrt{2\pi}} \int_{\partial L(t)} \frac{1}{\| \nabla r \|} \, d\mathcal{H}. \]

With similar arguments, we have
\[ C_1 \frac{\sqrt{2tK}}{\sqrt{2\pi}} \int_{\partial L(t)} \frac{d\mathcal{H}}{\| \nabla r \|} \leq \lim_{n \to \infty} \sqrt{nh^d} I_n \leq \lim_{n \to \infty} \sqrt{nh^d} I_n \leq C_2 \frac{\sqrt{2tK}}{\sqrt{2\pi}} \int_{\partial L(t)} \frac{d\mathcal{H}}{\| \nabla r \|}. \]

These inequalities, Lemma 3.4 and Lemma 3.3 give us, for all \( t \in \Theta \),
\[ \min \left( \lim_{n \to \infty} \sqrt{nh^d} \int_{V_n} P_{n,t}(x) \, dx, \lim_{n \to \infty} \sqrt{nh^d} \int_{V_n} \mathbb{P}_{n,t}(x) \, dx \right) \geq C_1 \frac{2tK}{\pi} \int_{\partial L(t)} \frac{d\mathcal{H}}{\| \nabla r \|}, \]
and
\[ \max \left( \lim_{n \to \infty} \sqrt{nh^d} \int_{V_n} P_{n,t}(x) \, dx, \lim_{n \to \infty} \sqrt{nh^d} \int_{V_n} \mathbb{P}_{n,t}(x) \, dx \right) \leq C_2 \frac{2tK}{\pi} \int_{\partial L(t)} \frac{d\mathcal{H}}{\| \nabla r \|}. \]

and Lemma 3.1 concludes the proof. \qed

References


